

Ten Years Back, Five Years Forward: The Data Seal of Approval

Ingrid Dillo

Data Archiving and Networked Services
(DANS)

Lisa de Leeuw

Data Archiving and Networked Services
(DANS)

Abstract

If we want to share data, the long-term storage of those data in a trustworthy digital archive is an essential condition. Trust is the basis of storing and sharing data. That trust must be present in the various stakeholders involved. Certification of digital archives can make an important contribution to the confidence of these stakeholders in the digital archives.

Ten years ago DANS was assigned the task of developing a Seal of Approval for digital data to ensure that archived data can still be found, understood and used in the future. In 2009 this Data Seal of Approval (DSA) was transferred to an international body, the DSA Board, which has managed and further developed the guidelines and the peer review process ever since.

The objectives of the DSA are to safeguard data, ensure high quality and guide reliable management of data for the future without requiring implementation of new standards, regulations or heavy investments. The DSA contains 16 guidelines for applying and verifying quality aspects concerning the creation, storage, use and reuse of digital data.

Based on feedback from data archives that applied for a DSA and different case studies we have gained some insight into the benefits of DSA. Still, the impact of having the Seal is not easy to measure. Seal holders usually refer to qualitative benefits in the form of increased awareness of the value of their repositories to their communities, funders and publishers.

Ten years down the line we can safely state that the Data Seal of Approval has proven its added value. If we try to look five years into the future, what can we expect? There are different developments: a growing interest in DSA among European research infrastructures, the collaboration between DSA and the ISCU World Data System under the umbrella of the RDA (Research Data Alliance) and the European Commission is showing a growing interest in certification services.

The success of DSA also provides the challenge to further professionalize the DSA organization in the coming years, this to enable its community to continue to grow. All in all there are promising developments for a bright future for the Data Seal of Approval.

Received 13 January 2015 | *Accepted* 10 February 2015

Correspondence should be addressed to Ingrid Dillo, DANS, PO Box 93067, 2509 AB The Hague, The Netherlands.
Email: ingrid.dillo@dans.knaw.nl

An earlier version of this paper was presented at the 10th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



Background and History

If we want to share data, the long-term storage of those data in a trustworthy digital archive is a *sine qua non*. Data created and used by scientists should be managed, curated and archived in order to preserve the initial investment in collecting them. Researchers must be certain that the data provided by the archives remain useful and meaningful, even in the long term. In addition, the archives should have sustainable business models themselves.

The concept of sustainability involves many challenging aspects in many areas: organizational, technical, financial, legal, etc. Certification can be an important contribution to ensuring the reliability and durability of digital archives, and hence the possibilities for sharing data over a long term.

Trust is the basis of storing and sharing data. That trust must be present in the various stakeholders. Data depositors want the assurance that their data in the digital archive are safe and will remain accessible, usable and meaningful. Data users have questions like: Have the data been well kept? Have they retained their authenticity and integrity? Are the data of good quality? Do the identifiers refer to the appropriate objects? Funders have other concerns. They want to be certain that their investment in data production yields optimum returns, i.e. that the data will be available for long-term reuse.

What characteristics make digital archives reliable? First, a digital archive's mission should be to give reliable, long-term access to the digital data under their care, now and in the future. Second, there should be permanent monitoring, planning and maintenance. The threats and risks within their systems must be understood. Finally, there should be a regular audit and certification cycle in place. Reliability is not something that, once achieved, can then be taken for granted.

Certification of digital archives can make an important contribution to the confidence of various stakeholders in the digital archives.

Ten years ago, in 2005, DANS¹ (Data Archiving and Networked Services) was established by the two main Dutch science organizations, KNAW² and NWO³. The mission of DANS is to promote sustained access to digital research data.

The two founding organisations assigned DANS the task of developing a Seal of Approval for digital data to ensure that archived data can still be found, understood and used in the future. A few years later the first edition of *Data Seal of Approval: Quality guidelines for digital research data* was presented at an international conference. The seal was initially developed for use in the Netherlands, but it was soon found to be very useful in an international context too. In 2009 the Data Seal of Approval was therefore transferred to an international body, the DSA Board, which has managed and further developed the guidelines and the peer review process ever since.

1 DANS: <http://www.dans.knaw.nl>

2 KNAW: <http://www.knaw.nl>

3 NWO: <http://www.nwo.nl>

The Data Seal of Approval

The objectives of the Data Seal of Approval are to safeguard data, to ensure high quality and to guide reliable management of data for the future without requiring the implementation of new standards, regulations or heavy investments.

The Data Seal of Approval:

- Gives researchers the assurance that their data will be stored in a reliable manner and can be reused;
- Provides funding bodies with the confidence that research data will remain available for reuse;
- Enables researchers to assess the repositories that hold the data which they want to reuse in a reliable manner;
- Supports data repositories in the efficient archiving and distribution of data.

The Guidelines

The Data Seal of Approval involves 16 guidelines for applying and verifying quality aspects concerning the creation, storage, use and reuse of digital data (DSA, 2014). The guidelines have been designed with a focus on scientific materials, but they can be applied to all types of digital information. The guidelines serve as the basis for awarding the Data Seal of Approval by the DSA Board.

The criteria for awarding the Data Seal of Approval to data repositories are in accordance with national and international guidelines for digital data archiving, such as the Kriterienkatalog vertrauenswürdige digitale Langzeitarchive developed by NESTOR, the Digital Repository Audit Method Based on Risk Assessment (DRAMBORA)⁴ published by the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE), and Trustworthy Repositories Audit and Certification (TRAC)⁵: Criteria and Checklist of the Research Library Group (RLG). The following publications have also been taken into account: Foundations of Modern Language Resource Archives by the Max Planck Institute (Wittenburg et al., 2006), and Stewardship of Digital Research Data: A Framework of Principles and Guidelines by the Research Information Network (RIN, n.d.). The DSA guidelines can be seen as a minimum set distilled from the above proposals.

Fundamental to the guidelines are five principles that together determine whether or not the digital data may be considered as sustainably archived:

- The data can be found on the Internet.
- The data are accessible, while taking into account relevant legislation with regard to personal information and intellectual property.
- The data are available in a usable format.
- The data are reliable.
- The data can be referred to (persistent identifiers).

⁴ DRAMBORA: <http://www.dcc.ac.uk/resources/repository-audit-and-assessment/drambora>

⁵ TRAC: <http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying>

These principles are integral to the guidelines, which focus on three stakeholders:

- The data producer, who is responsible for the quality of the digital data;
- The data repository, who is responsible for the quality of storage and availability of the data (data management);
- The data consumer, who is responsible for the quality of use of the data.

The basic assumption is that the data repository is responsible for enabling and supporting data producers' and data consumers' compliance with the guidelines. A data repository is designated a Trusted Digital Repository (TDR) if it complies with Guidelines 4 to 13 and if it enables data producers and data consumers to comply with Guidelines 1 to 3 and 14 to 16.

Guidelines for data producers

The quality of the digital research data is determined by:

- Their intrinsic value to their sector (designated community): scientific, scholarly, business, etc.;
- The format in which the data and supporting information are stored;
- The documentation (metadata or contextual information) supporting the data.

Therefore, the data producer deposits the data with sufficient information, in the recommended format and with the requested metadata.

Guidelines for data repositories

The data repository is responsible for access and preservation of digital data in the long term. Two factors, in particular, determine the quality of the data repository:

- The quality of the organizational framework in which the data repository is incorporated (organization and processes);
- The quality of the technical infrastructure of the data repository.

Organizations that play a role in digital archiving and are establishing a Trusted Digital Repository shall possess a sound financial, organizational and legal basis in the long term.

Guidelines for data consumers

The data consumer uses the digital data in compliance with the relevant guidelines, dealing with access regulations, licenses and codes of conduct.

The Procedures

The starting point for obtaining the Data Seal of Approval is the website,⁶ where an application form can be submitted. Once the DSA Board receives the form, a self-assessment is made available in the DSA online tool. The self-assessment is meant to supply evidence that the applicant data repository meets the 16 DSA guidelines and the relevant level of compliance. A description of the context of the data repository is also required.

⁶ DSA: <http://www.datasealofapproval.org/>

After the submission of the self-assessment by the data repository, the DSA Board appoints a peer reviewer who is given a two months' time frame in which to evaluate the self-assessment. The peer reviewer will either confirm the evidence or require additional information depending on the adherence to the guidelines and the level of compliance. Resubmission of the modified application and requests for additional information by the peer reviewer will continue until the reviewer is satisfied with the evidence and awards the DSA. In the event of a dispute, the applicant data repository can contact the DSA Board.

As long as a self-assessment is in the application process, it will not be made public. The self-assessment, including all evidence, will only be published on the websites of the DSA and the applicant data repository after the DSA has been awarded.

After the Data Seal of Approval is awarded by the DSA Board, the DSA logo may be displayed on the repository's website. At the same time, the DSA Board will post the approved assessment, including evidence and peer review comments, on the DSA website.

A Data Seal of Approval for a given period can be displayed indefinitely but will need to be updated periodically if the repository wants to stay compliant with newly released guidelines and receive the latest DSA logo. DSA-certified repositories will be contacted automatically when an update is available.

The DSA Community

The Data Seal of Approval is driven by the voluntary involvement of all stakeholders. The organization of the DSA is established by Regulations⁷, which are available on the DSA website. The Regulations define the various rights and duties of the DSA Community. The world of the DSA is made up of a number of components:

- The DSA Community comprises all of the organizations with one or more DSA-certified repositories.
- The DSA General Assembly is the governing body of the DSA Community. The General Assembly elects the DSA Board and provides the Board with advice when needed. General Assembly members commit to conducting a maximum of three peer reviews a year to ensure that the DSA remains community-driven and sustainable.
- The DSA Board is drawn from and elected by the General Assembly representatives. The Board conducts the daily business of the DSA Community, manages and monitors the DSA assessment procedure, convenes meetings of the General Assembly and informs the DSA Community about all DSA activities.
- Peer reviewers belong to one of the organizations in the General Assembly and have completed at least one self-assessment, which resulted in the award of the latest DSA. They review and assess evidence in a timely, complete and impartial manner, ensuring that DSA applications remain confidential until the DSA is awarded.

⁷ DSA Regulations: http://datasealofapproval.org/media/filer_public/2013/09/27/dsa-regulations_2013.pdf

Ten Years Back: Experiences and Lessons Learned

The idea to develop a basic seal of approval for digital archives originated in the Netherlands ten years ago. From that moment the Data Seal of Approval organically grew, slowly but surely, from a national into an international certification standard.

Next to the DSA a number of other certification standards have become available over the last few years.

The nestorSeal⁸ provides a second set of guidelines. The 34 criteria were developed by the German organization NESTOR (a consortium of museums, archives and libraries) and formalized as the DIN 31644 standard⁹. It is expected that the first nestorSeals will be awarded in 2015.

The third way to evaluate a digital archive is provided by ISO standard 16363¹⁰. The standard is very detailed and contains more than one hundred criteria for different aspects of a digital archive. They focus on organizational infrastructure, digital object management, and infrastructure and risk management. In 2011, six test audits were performed: three in Europe and three in the US. The ISO standard is based on a formal external audit of the archive, formalized in ISO 16919: Requirements for bodies providing audit and certification of candidate trustworthy digital repositories¹¹.

How do these standards fit together? In 2010, a Memorandum of Understanding¹² (MoU) was signed by the parties involved in these three standards. The purpose of the MoU was to set up a comprehensive multi-level framework for the certification of digital archives. This European Framework for Audit and Certification of Digital Repositories offers three evaluation levels of increasing reliability.

Basic certification is granted to repositories qualifying for the DSA. Extended certification is granted to repositories with Basic Certification that perform an additional structured, externally reviewed and publicly available self-audit based on ISO 16363 or DIN 31644. Finally, formal certification is granted to repositories which, in addition to Basic Certification, pass a full external audit and certification based on ISO 16363 or DIN 31644.

Although more options for certification have become available for digital archives, DSA is doing well. The DSA community is growing and thriving. The seal acquires more prestige as the number and geographical spread of seals grows. Today 37 Seals have been awarded and some 35 digital archives are working on their DSA self-assessment.¹³ This steady growth shows that there is a clear demand for a basic way to assess the trustworthiness of a digital archive.

Based on feedback from data archives that applied for a DSA and different case studies presented at the annual Data Seal of Approval (DSA) conferences we have gained some insight into the benefits of DSA. Still, the impact of having the Seal is not easy to measure. Seal holders usually refer to qualitative benefits in the form of increased awareness of the value of their repositories to their communities, funders and publishers.

⁸ nestorSeal: http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.html

⁹ DIN 31644: <http://www.nabd.din.de/cmd?level=tpl-art-detailansicht&committeeid=54738855&artid=147058907&languageid=en>

¹⁰ ISO 16363: http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510

¹¹ ISO 16919: http://www.iso.org/iso/catalogue_detail.htm?csnumber=57950

¹² Trusted Digital Repository (MoU):

http://datasealofapproval.org/media/filer_public/2014/08/28/20100709_020_signedmoutocreateeurop_eanframeworkforauditandcertificationofdigitalrepositories.pdf

¹³ See <http://datasealofapproval.org/en/assessment/> for the list of acquired seals.

What are these qualitative benefits of DSA?

- **Stakeholder confidence:** Having the Data Seal of Approval signifies to funders that the data they have invested in will continue to be available for reuse. Data producers can be confident that the data they have worked hard to create will be protected, and data consumers can be sure that the data they are using have been managed optimally.
- **Improvements in communication:** Preparing for the self-assessment prompts a repository to communicate internally about their overall mission and goals in ways not always present in day-to-day interactions.
- **Improvement in processes:** Conducting the self-assessment stimulates a repository to improve its processes and procedures and move to a higher level of professionalism, with an incentive to improve its operations over time.
- **Transparency:** The DSA is designed to provide an open statement of repository evidence enabling anyone to evaluate the repository's operations and policies.
- **Differentiation from others:** There are a growing number of options for depositing data. Having the DSA sets a repository apart from others and enhances its reputation, showing in an easily recognized way that the repository is following good practice.
- **Awareness raising about digital preservation:** In this age of instant communication, people often focus on access to digital resources but do not consider the importance of preserving data for future reuse. Complying with the 16 DSA guidelines shows a commitment to ensuring that data will remain usable for new generations.
- **Less labor- and time-intensive:** The 16 guidelines of the DSA are the entry level of the European Framework for Digital Certification, in contrast to the 34 criteria for DIN31644 or over 100 metrics in ISO16363. There is no site visit as the assessment is conducted online through an efficient tool.

Jenny Mitcham and Catherine Hardman (2011) of the UK Archaeology Data Service (ADS) have indicated the two primary reasons to apply for the DSA in their DSA case study¹⁴. They wanted to reflect on their own performance and also wanted to be able to demonstrate to their peers and user base that ADS is a trustworthy repository for their data. Furthermore, they are convinced that gaining the DSA embeds ADS within a community of archives working to higher standards and potentially allows the archive to benefit from closer ties and relationships with them. It opens up possibilities of working with others to enhance ADS policies and procedures.

Of course there are also challenges facing the Data Seal of Approval. The DSA standard has grown in an organic way and has always taken a community-based approach. There is no legal entity and the seal is neither a formally deposited nor a protected certification standard. DSA activities have always taken place on a voluntary basis.

Because of this the DSA community has proven to be agile, flexible and pragmatic and the application of a seal has always been free of charge. On the other hand this voluntary basis and the lack of an earmarked budget hampers the development of a

¹⁴ For more case studies see DSA references for completing your self-assessment: <http://datasealofapproval.org/en/assessment/> and EUDAT: <http://www.eudat.eu/events/certification>

more consistent and firm reviewing process. The aim is to have at least two peer reviewers for each self-assessment and to train these reviewers, in order to further improve consistency across reviews.

Acquiring the DSA will not immediately lead to a steep rise in the number of deposits and the re-use of the data of a digital archive. Showing the DSA-logo on the website of a digital archive should be a sign for data depositors that the data they have worked hard to create will be protected, and for data consumers that they can be sure that the data they are using have been managed optimally. But in practice, researchers looking to deposit or re-use data are obviously more focused on the reputation of a digital archive within their research community than on the presence of a DSA-logo.

The added value of the DSA is more strongly related to gaining trust among funders and publishers. National and international funders are increasingly demanding open data and data management policies that implicate the long term storage and accessibility of (a selection of) data. The DSA helps funders to identify trustworthy digital archives to which they can refer the researchers they fund. The DSA can also help publishers with a data availability policy to refer their authors to trustworthy digital archives where they can safely store the data underlying their articles.

Five Years Forward: The Future of DSA

Ten years down the line we can safely state that the Data Seal of Approval has proven its added value. If we try to look five years forward, what can we expect? The following developments are taking place.

First of all we see a growing interest in DSA among European research infrastructures. Within these infrastructures, building confidence in the services offered is considered increasingly important. In this context, infrastructures such as CESSDA¹⁵, CLARIN¹⁶ and DARIAH¹⁷ are looking at the DSA guidelines.

CLARIN has already made DSA certification mandatory for all its centres. CESSDA is working to integrate the DSA guidelines with their own infrastructure and DARIAH is considering the adoption of the guidelines. In the proposal for the continuation of the European EUDAT¹⁸ project, DSA also plays a significant role. This development will most probably lead to a growth of the number of seals acquired in Europe in the coming years.

A second development that is in progress is the collaboration between DSA and the ISCU World Data System¹⁹ under the umbrella of the RDA (Research Data Alliance)²⁰.

Recently the RDA Working Group Repository Audit and Certification: DSA-WDS Partnership²¹ was launched. In this group, the DSA Board collaborates with the scientific committee of ICSU/WDS. The World Data System (WDS) is a body of the International Council for Science (ICSU), whose data archives can be members. The WDS requires some categories of membership to go through an accreditation process.

¹⁵ CESSDA: <http://www.cessda.net/>

¹⁶ CLARIN: <http://clarin.eu/>

¹⁷ DARIAH: <http://www.dariah.eu/>

¹⁸ EUDAT: <http://eudat.eu/>

¹⁹ ISCU World Data System: <https://www.icsu-wds.org/>

²⁰ RDA: <https://rd-alliance.org/node>

²¹ RDA WG Repository Audit and Certification – DSA-WDS Partnership: <https://rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93wds-partnership-wg.html>

The accreditation criteria are very similar to DSA. Between DSA and WDS there is a complementarity in geographical spread and in scientific disciplines. The experience of DSA lies mainly within the social sciences and humanities, while the WDS has a strong focus on the earth and space sciences. It was therefore decided to explore the possibilities of collaboration in this working group. Although it is not clear yet what the concrete outcomes of this working group will be, collaboration could lead to more efficiency and more certifications in the future.

The third development surrounds the Recommendation on Access to and Preservation of Scientific Information published by the European Commission (2012). In this recommendation, the Commission encouraged a European open access policy. Last year this was followed by a limited pilot action on open access to research data in Horizon 2020 (European Commission, 2013). Participating projects are required to develop a Data Management Plan in which they will specify what data will be open. In this DMP the researcher also needs to specify in which repository the data will be stored and how the long term preservation and accessibility of these data will be guaranteed. In future the certification of digital archives will most probably play a role here as well.

A first indication of this is the paragraph on certification services within the Horizon 2020 e-Infrastructure call of 2014-2015²². In this paragraph the Commission solicits for proposals aimed at providing ‘services to ensure the quality and reliability of the e-infrastructure, including certification mechanisms for repositories and certification services to test and benchmark capabilities in terms of resilience and service continuity of e-infrastructures.’

Within this call a proposal has been submitted by the main stakeholders of the four certification standards: DSA, DIN, ISO and ICSU/WDS. This CTRUST proposal consists of two strands. The first one is aimed at the further professionalization of and collaboration between the four standards. This includes the development of an integrated framework for certification of trustworthy digital repositories, online tools for risk analysis and assessment of repositories, consultancy for digital repositories, the development of sustainable business models for permanent certification services, the availability of scalable online tools and, last but not least, the training of reviewers and auditors. The second strand aims to boost the number of European trustworthy repositories. The aim is to certify over 60 repositories. The candidate repositories will be selected by an independent selection committee and will receive a limited financial incentive. Synergy with the work within research communities will be established through collaboration with existing European projects, infrastructures and organizations.

The success of DSA provides the challenge to further professionalize the DSA organization in the coming years in order to enable its community to continue to grow. If this proposal receives funding it will provide an enormous boost and enable DSA to transform into a sustainable and professional certification service. If not, DSA will continue to develop at a slower pace, with a primary focus on strengthening its community and further improving the consistency of the review process.

References

DSA. (2014). *Data Seal of Approval guidelines 2014-2015*. Retrieved from <http://datasealofapproval.org/en/information/guidelines/>

²² H2020-EINFRA-2014-2: <http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/2137-einfra-1-2014.html>

- European Commission. (2012). *Recommendation on access to and preservation of scientific information*. Retrieved from https://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf
- European Commission. (2013). *Guidelines on Data Management in H2020. Version 1.0*. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Mitcham, J., & Hardman, C. (2011). *ADS and the Data Seal of Approval: Case study for the DCC*. Retrieved from <http://www.dcc.ac.uk/resources/case-studies/ads-dsa>
- RIN. (n.d.). *Stewardship of digital research data: A framework of principles and guidelines*. Retrieved from <http://www.webarchive.org.uk/wayback/archive/20120107235537/http://www.rin.ac.uk/our-work/data-management-and-curation/stewardship-digital-research-data-principles-and-guidelines>
- Wittenburg, P., Broeder, D., Klein, W., Levinson, S., & Romary, L. (2006). *Foundations of modern language resource archives*. Retrieved from http://pubman.mpdl.mpg.de/pubman/item/escidoc:58934:4/component/escidoc:58935/Wittenburg_2006_foundations.pdf